# RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.**

Application Serial Number: 10/089,027A

Source: IF W16

Date Processed by STIC: 8-20-05

# *ENTERED*

IFW16

RAW SEQUENCE LISTING                    DATE: 08/20/2005
PATENT APPLICATION:  US/10/089,027A      TIME: 11:08:31

Input Set : A:\NEB-166-PUS.APP.txt
Output Set: N:\CRF4\08202005\J089027A.raw

```
 3 <110> APPLICANT: JACK, WILLIAM E.
 4       GARDNER, ANDREW
 5       BUZBY, PHILIP R.
 6       DiMEO, JAMES J.
 7       NEW ENGLAND BIOLABS, INC.
 8       NEN LIFE SCIENCE PRODUCTS, INC.
10 <120> TITLE OF INVENTION: INCORPORATION OF MODIFIED NUCLEOTIDES BY ARCHAEON DNA
11       POLYMERASES AND RELATED METHODS
13 <130> FILE REFERENCE: NEB-166-PUS
15 <140> CURRENT APPLICATION NUMBER: 10/089,027A
16 <141> CURRENT FILING DATE: 2002-03-26
18 <150> PRIOR APPLICATION NUMBER: PCT/US00/26900
19 <151> PRIOR FILING DATE: 2000-09-29
21 <150> PRIOR APPLICATION NUMBER: 60/157,204
22 <151> PRIOR FILING DATE: 1999-09-30
24 <160> NUMBER OF SEQ ID NOS: 33
26 <170> SOFTWARE: PatentIn Ver. 2.0
28 <210> SEQ ID NO: 1
29 <211> LENGTH: 36
30 <212> TYPE: DNA
31 <213> ORGANISM: Artificial Sequence
33 <220> FEATURE:
34 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
35       oligonucleotide
37 <400> SEQUENCE: 1
38 caggcagagg cttataaaaa tcctcgccaa cagctt                    36
40 <210> SEQ ID NO: 2
41 <211> LENGTH: 26
42 <212> TYPE: DNA
43 <213> ORGANISM: Artificial Sequence
45 <220> FEATURE:
46 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
47       oligonucleotide
49 <400> SEQUENCE: 2
50 ggtggcagca gccaactcag cttcct                    26
52 <210> SEQ ID NO: 3
53 <211> LENGTH: 24
54 <212> TYPE: DNA
55 <213> ORGANISM: Artificial Sequence
57 <220> FEATURE:
58 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
59       oligonucleotide
61 <400> SEQUENCE: 3
```

```
62 gattctcatg ataagctacg ccga                                         24
64 <210> SEQ ID NO: 4
65 <211> LENGTH: 5837
66 <212> TYPE: DNA
67 <213> ORGANISM: Thermococcus litoralis
69 <400> SEQUENCE: 4
70 gaattcgcga taaaatctat tttcttcctc catttttcaa tttcaaaaac gtaagcatga 60
71 gccaaacctc tcgccctttc tctgtccttc ccgctaaccc tcttgaaaac tctctccaaa 120
72 gcatttttg atgaaagctc acgctcctct atgagggtca gtatatctgc aatgagttcg 180
73 tgaagggtta ttctgtagaa caactccatg attttcgatt tggatggggg tttaaaaatt 240
74 tggcggaact tttatttaat ttgaactcca gtttatatct ggtggtattt atgatactgg 300
75 acactgatta cataacaaaa gatggcaagc ctataatccg aatttttaag aaagagaacg 360
76 gggagtttaa aatagaactt gaccctcatt ttcagcccta tatatatgct cttctcaaag 420
77 atgactccgc tattgaggag ataaaggcaa taaagggcga gagacatgga aaaactgtga 480
78 gagtgctcga tgcagtgaaa gtcaggaaaa aattttgggg aagggaagtt gaagtctgga 540
79 agctcatttt cgagcatccc caagacgttc cagctatgcg gggcaaaata agggaacatc 600
80 cagctgtggt tgacatttac gaatatgaca tacccttgc caagcgttat ctcatagaca 660
81 agggcttgat tcccatggag ggagacgagg agcttaagct ccttgccttt gatattgaaa 720
82 cgttttatca tgagggagat gaatttggaa agggcgagat aataatgatt agttatgccg 780
83 atgaagaaga ggccagagta atcacatgga aaaatatcga tttgccgtat gtcgatgttg 840
84 tgtccaatga aagagaaatg ataaagcgtt ttgttcaagt tgttaaagaa aaagaccccg 900
85 atgtgataat aacttacaat ggggacaatt ttgatttgcc gtatctcata aaacgggcag 960
86 aaaagctggg agttcggctt gtcttaggaa gggacaaaga acatcccgaa cccaagattc 1020
87 agaggatggg tgatagtttt gctgtggaaa tcaagggtag aatccacttt gatctttcc 1080
88 cagttgtgcg aaggacgata aacctcccaa cgtatacgct tgaggcagtt tatgaagcag 1140
89 ttttaggaaa aaccaaaagc aaattaggag cagaggaaat tgccgctata tgggaaacag 1200
90 aagaaagcat gaaaaaacta gcccagtact caatggaaga tgctaggca acgtatgagc 1260
91 tcgggaagga attcttcccc atggaagctg agctggcaaa gctgataggt caaagtgtat 1320
92 gggacgtctc gagatcaagc accggcaacc tcgtggagtg gtatcttta agggtggcat 1380
93 acgcgaggaa tgaacttgca ccgaacaaac ctgatgagga agagtataaa cggcgcttaa 1440
94 gaacaactta cctgggagga tatgtaaaag agccagaaaa aggtttgtgg gaaaatatca 1500
95 tttatttgga tttccgcagt ctgtaccctt caataatagt tactcacaac gtatccccag 1560
96 ataccccttga aaaagagggc tgtaagaatt acgatgttgc tccgatagta ggatataggt 1620
97 tctgcaagga ctttccgggc tttattccct ccatactcgg ggacttaatt gcaatgaggc 1680
98 aagatataaa gaagaaaatg aaatccacaa ttgacccgat cgaaaagaaa atgctcgatt 1740
99 ataggcaaag ggctattaaa ttgcttgcaa acagcatctt acccaacgag tggttaccaa 1800
100 taattgaaaa tggagaaata aaaattcgtga aaattggcga gtttataaac tcttacatgg 1860
101 aaaaacagaa ggaaaacgtt aaaacagtag agaatactga agttctcgaa gtaaacaacc 1920
102 tttttgcatt ctcattcaac aaaaaaatca agaaagtga agtcaaaaaa gtcaaagccc 1980
103 tcataagaca taagtataaa gggaaagctt atgagattca gcttagctct ggtagaaaaa 2040
104 ttaacataac tgctggccat agtctgttta cagttagaaa tggagaaata aaggaagttt 2100
105 ctggagatgg gataaaagaa ggtgacctta ttgtagcacc aaagaaaatt aaactcaatg 2160
106 aaaaaggggt aagcataaac attcccgagt taatctcaga tctttccgag gaagaaacag 2220
107 ccgacattgt gatgacgatt tcagccaagg gcagaaagaa cttctttaaa ggaatgctga 2280
108 gaactttaag gtggatgttt ggagaagaaa atagaaggat aagaacattt aatcgctatt 2340
109 tgttccatct cgaaaaacta ggccttatca aactactgcc ccgcggatat gaagttactg 2400
110 actgggagag attaaagaaa tataaacaac tttacgagaa gcttgctgga agcgttaagt 2460
111 acaacggaaa caagagagag tatttagtaa tgttcaacga gatcaaggat tttatatctt 2520
112 acttcccaca aaaagagctc gaagaatgga aaattggaac tctcaatggc tttagaacga 2580
```

```
113 attgtattct caaagtcgat gaggattttg ggaagctcct aggttactat gttagtgagg 2640
114 gctatgcagg tgcacaaaaa aataaaactg gtggtatcag ttattcggtg aagctttaca 2700
115 atgaggaccc taatgttctt gagagcatga aaaatgttgc agaaaaattc tttggcaagg 2760
116 ttagagttga cagaaattgc gtaagtatat caaagaagat ggcatactta gttatgaaat 2820
117 gcctctgtgg agcattagcc gaaaacaaga gaattccttc tgttatactc acctctcccg 2880
118 aaccggtacg gtggtcattt ttagaggcgt attttacagg cgatggagat atacatccat 2940
119 caaaaaggtt taggctctca acaaaaagcg agctccttgc aaatcagctt gtgttcttgc 3000
120 tgaactcttt gggaatatcc tctgtaaaga taggctttga cagtggggtc tatagagtgt 3060
121 atataaatga agacctgcaa tttccacaaa cgtctaggga gaaaaacaca tactactcta 3120
122 acttaattcc caaagagatc cttagggacg tgtttggaaa agagttccaa aagaacatga 3180
123 cgttcaagaa atttaaagag cttgttgact ctggaaaact aacagggag aaagccaagc 3240
124 tcttggagtt cttcattaat ggagatattg tccttgacag agtcaaaagt gttaaagaaa 3300
125 aggactatga agggtatgtc tatgacctaa gcgttgagga taacgagaac tttcttgttg 3360
126 gttttggttt gctctatgct cacaacagct attacggcta tatggggtat cctaaggcaa 3420
127 gatggtactc gaaggaatgt gctgaaagcg ttaccgcatg ggggagacac tacatagaga 3480
128 tgacgataag agaaatagag gaaaagttcg gctttaaggt tctttatgcg gacagtgtct 3540
129 caggagaaag tgagatcata ataaggcaaa acggaaagat tagatttgtg aaaataaagg 3600
130 atcttttctc taaggtggac tacagcattg gcgaaaaaga atactgcatt ctcgaaggtg 3660
131 ttgaagcact aactctggac gatgacggaa agcttgtctg gaagcccgtc ccctacgtga 3720
132 tgaggcacag agcgaataaa agaatgttcc gcatctggct gaccaacagc tggtatatag 3780
133 atgttactga ggatcattct ctcataggct atctaaacac gtcaaaaacg aaaactgcca 3840
134 aaaaaatcgg ggaaagacta aaggaagtaa agcctttga attaggcaaa gcagtaaaat 3900
135 cgctcatatg cccaaatgca ccgttaaagg atgagaatac caaaactagc gaaatagcag 3960
136 taaaattctg ggagctcgta ggattgattg taggagatgg aaactggggt ggagattctc 4020
137 gttgggcaga gtattatctt ggactttcaa caggcaaaga tgcagaagag ataaagcaaa 4080
138 aacttctgga acccctaaaa acttatggag taatctcaaa ctattaccca aaaaacgaga 4140
139 aaggggactt caacatcttg gcaaagagcc ttgtaaagtt tatgaaaagg cactttaagg 4200
140 acgaaaaagg aagacgaaaa attccagagt tcatgtatga gcttccggtt acttacatag 4260
141 aggcatttct acgaggactg ttttcagctg atggtactgt aactatcagg aagggagttc 4320
142 cagagatcag gctaacaaac attgatgctg actttctaag ggaagtaagg aagcttctgt 4380
143 ggattgttgg aatttcaaat tcaatatttg ctgagactac tccaaatcgc tacaatggtg 4440
144 tttctactgg aacctactca aagcatctaa ggatcaaaaa taagtggcgt tttgctgaaa 4500
145 ggataggctt tttaatcgag agaaagcaga agagactttt agaacattta aaatcagcga 4560
146 gggtaaaaag gaataccata gattttggct ttgatcttgt gcatgtgaaa aaagtcgaag 4620
147 agataccata cgagggttac gtttatgaca ttgaagtcga agagacgcat aggttctttg 4680
148 caaacaacat cctggtacac aatactgacg gcttttatgc cacaataccc ggggaaaagc 4740
149 ctgaactcat taaaaagaaa gccaaggaat tcctaaacta cataaactcc aaacttccag 4800
150 gtctgcttga gcttgagtat gagggctttt acttgagagg attctttgtt acaaaaaagc 4860
151 gctatgcagt catagatgaa gagggcagga taacaacaag gggcttggaa gtagtaagga 4920
152 gagattggag tgagatagct aaggagactc aggcaaaggt tttagaggct atacttaaag 4980
153 aggaagtgt tgaaaaagct gtagaagttg ttagagatgt tgtagagaaa atagcaaaat 5040
154 acaggggttcc acttgaaaag cttgttatcc atgagcagat taccagggat ttaaaggact 5100
155 acaaagccat tggccctcat gtcgcgatag caaaaagact tgccgcaaga gggataaaag 5160
156 tgaaaccggg cacaataata agctatatcg ttctcaaagg gagcggaaag ataagcgata 5220
157 gggtaatttt acttacagaa tacgatccta gaaaacacaa gtacgatccg gactactaca 5280
158 tagaaaacca agttttgccg gcagtactta ggatactcga agcgtttgga tacagaaagg 5340
159 aggatttaag gtatcaaagc tcaaaacaaa ccggcttaga tgcatggctc aagaggtagc 5400
160 tctgttgctt tttagtccaa gtttctccgc gagtctctct atctctcttt tgtattctgc 5460
161 tatgtggttt tcattcacta ttaagtagtc cgccaaagcc ataacgcttc caattccaaa 5520
```

```
162 cttgagctct ttccagtctc tggcctcaaa ttcactccat gttttggat cgtcgcttct 5580
163 ccctcttctg ctaagcctct cgaatctttt tcttggcgaa gagtgtacag ctatgatgat 5640
164 tatctcttcc tctggaaacg catctttaaa cgtctgaatt tcatctagag acctcactcc 5700
165 gtcgattata actgccttgt acttctttag tagttctttt acctttggga tcgttaattt 5760
166 tgccacggca ttgtccccaa gctcctgcct aagctgaatg ctcacactgt tcataccttc 5820
167 gggagttctt gggatcc                                             5837
169 <210> SEQ ID NO: 5
170 <211> LENGTH: 15
171 <212> TYPE: PRT
172 <213> ORGANISM: Thermococcus litoralis
174 <400> SEQUENCE: 5
175 Ala Ile Lys Leu Leu Ala Asn Ser Tyr Tyr Gly Tyr Met Gly Tyr
176  1               5                10                15
179 <210> SEQ ID NO: 6
180 <211> LENGTH: 15
181 <212> TYPE: PRT
182 <213> ORGANISM: Pyrococcus Sp. (GB-D)
184 <400> SEQUENCE: 6
185 Ala Ile Lys Ile Leu Ala Asn Ser Tyr Tyr Gly Tyr Tyr Gly Tyr
186  1               5                10                15
189 <210> SEQ ID NO: 7
190 <211> LENGTH: 15
191 <212> TYPE: PRT
192 <213> ORGANISM: Thermococcus sp.
194 <400> SEQUENCE: 7
195 Ala Ile Lys Ile Leu Ala Asn Ser Phe Tyr Gly Tyr Tyr Gly Tyr
196  1               5                10                15
199 <210> SEQ ID NO: 8
200 <211> LENGTH: 15
201 <212> TYPE: PRT
202 <213> ORGANISM: Pyrococcus furiosus
204 <400> SEQUENCE: 8
205 Ala Ile Lys Leu Leu Ala Asn Ser Phe Tyr Gly Tyr Tyr Gly Tyr
206  1               5                10                15
209 <210> SEQ ID NO: 9
210 <211> LENGTH: 15
211 <212> TYPE: PRT
212 <213> ORGANISM: Thermococcus fumicolans
214 <400> SEQUENCE: 9
215 Ala Ile Lys Ile Leu Ala Asn Ser Phe Tyr Gly Tyr Tyr Gly Tyr
216  1               5                10                15
219 <210> SEQ ID NO: 10
220 <211> LENGTH: 15
221 <212> TYPE: PRT
222 <213> ORGANISM: Thermococcus gorgonarius
224 <400> SEQUENCE: 10
225 Ala Ile Lys Ile Leu Ala Asn Ser Phe Tyr Gly Tyr Tyr Gly Tyr
226  1               5                10                15
229 <210> SEQ ID NO: 11
```

```
230 <211> LENGTH: 15
231 <212> TYPE: PRT
232 <213> ORGANISM: Thermococcus sp. (TY)
234 <400> SEQUENCE: 11
235 Ala Val Lys Leu Leu Ala Asn Ser Tyr Tyr Gly Tyr Met Gly Tyr
236   1           5               10              15
239 <210> SEQ ID NO: 12
240 <211> LENGTH: 15
241 <212> TYPE: PRT
242 <213> ORGANISM: Pyrococcus abyssi
244 <400> SEQUENCE: 12
245 Ala Ile Lys Ile Leu Ala Asn Ser Tyr Tyr Gly Tyr Tyr Gly Tyr
246   1           5               10              15
249 <210> SEQ ID NO: 13
250 <211> LENGTH: 15
251 <212> TYPE: PRT
252 <213> ORGANISM: Pyrococcus glycovaorans
254 <400> SEQUENCE: 13
255 Ala Ile Lys Ile Leu Ala Asn Ser Tyr Tyr Gly Tyr Tyr Gly Tyr
256   1           5               10              15
259 <210> SEQ ID NO: 14
260 <211> LENGTH: 15
261 <212> TYPE: PRT
262 <213> ORGANISM: Pyrococcus horikoshii
264 <400> SEQUENCE: 14
265 Ala Ile Lys Ile Leu Ala Asn Ser Tyr Tyr Gly Tyr Tyr Gly Tyr
266   1           5               10              15
269 <210> SEQ ID NO: 15
270 <211> LENGTH: 15
271 <212> TYPE: PRT
272 <213> ORGANISM: Pyrococcus sp. (GE23)
274 <400> SEQUENCE: 15
275 Ala Ile Lys Ile Leu Ala Asn Ser Tyr Tyr Gly Tyr Tyr Gly Tyr
276   1           5               10              15
279 <210> SEQ ID NO: 16
280 <211> LENGTH: 15
281 <212> TYPE: PRT
282 <213> ORGANISM: Pyrococcus Sp. (KOD1)
284 <400> SEQUENCE: 16
285 Ala Ile Lys Ile Leu Ala Asn Ser Tyr Tyr Gly Tyr Tyr Gly Tyr
286   1           5               10              15
289 <210> SEQ ID NO: 17
290 <211> LENGTH: 15
291 <212> TYPE: PRT
292 <213> ORGANISM: Pyrococcus woesei
294 <400> SEQUENCE: 17
295 Ala Ile Lys Leu Leu Ala Asn Ser Phe Tyr Gly Tyr Tyr Gly Tyr
296   1           5               10              15
299 <210> SEQ ID NO: 18
```

RAW SEQUENCE LISTING ERROR SUMMARY     DATE: 08/20/2005
PATENT APPLICATION:  **US/10/089,027A**     TIME: 11:08:32

Input Set : **A:\NEB-166-PUS.APP.txt**
Output Set: **N:\CRF4\08202005\J089027A.raw**

**Please Note:**
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:31; N Pos. 2,3,11,12,14,46,77,84,128,152
Seq#:32; N Pos. 2,3,6,21,468,532,535,560,577
Seq#:33; N Pos. 8,9,11,20,26,27,31,35,40,46,60,68,75,119,124,132,139,143
Seq#:33; N Pos. 173,174,194,196,210,213,220,246,252,297,326,329,342,349,365
Seq#:33; N Pos. 368,378,419,428,435,454,502,537,551,571,597

**VERIFICATION SUMMARY**                     DATE: 08/20/2005
PATENT APPLICATION:   **US/10/089,027A**        TIME: 11:08:32

Input Set : **A:\NEB-166-PUS.APP.txt**
Output Set: **N:\CRF4\08202005\J089027A.raw**

```
L:556 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:31 after pos.:0
M:341 Repeated in SeqNo=31
L:626 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:32 after pos.:0
M:341 Repeated in SeqNo=32
L:861 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:33 after pos.:0
M:341 Repeated in SeqNo=33
```